

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Salah satu komponen utama yang menentukan kemajuan suatu negara adalah pendidikan. Pemerintah Indonesia telah mengambil beberapa kebijakan untuk meningkatkan akses dan kualitas pendidikan di berbagai jenjang, dari sekolah dasar hingga perguruan tinggi. Namun, perbedaan dalam tingkat pendidikan di berbagai provinsi menunjukkan adanya disparitas, atau perbedaan, yang perlu diteliti secara menyeluruh untuk mengidentifikasi penyebabnya dan menemukan solusi yang tepat.

Di era internet saat ini, teknik *data mining* untuk analisis pendidikan menjadi semakin penting. Algoritma pengelompokan *K-Means Clustering*, yang memungkinkan pengelompokan data berdasarkan atribut tertentu, adalah salah satu teknik yang dapat digunakan. Proses pengelompokan ini dapat dilakukan secara efisien dan efektif dengan menggunakan alat seperti *RapidMiner*, yang menyediakan *platform* untuk analisis data yang komprehensif.

Dalam analisis tingkat penyelesaian pendidikan, penerapan *K-Means Clustering* bertujuan untuk menemukan pola dan tren yang mungkin tersembunyi dalam data. Misalnya, dengan mengelompokkan data berdasarkan provinsi dan jenjang pendidikan, kita dapat mengetahui provinsi mana yang memiliki tingkat penyelesaian pendidikan yang tinggi dan rendah, serta faktor-faktor apa saja yang mungkin berkontribusi terhadap hal tersebut.

Penelitian sebelumnya, yang dilakukan oleh (Indriyani & Irfiani, 2019) berjudul *Clustering Data Penjualan Pada Toko Perlengkapan Outdoor* menggunakan Metode *K-Means Clustering*, mengungkapkan bahwa algoritma *K-Means Clustering* memiliki kemampuan untuk meningkatkan kepuasan konsumen dengan mengelompokkan data persediaan produk penjualan. Toko Genta Corp mengalami kesulitan dalam menentukan jumlah produk yang tersedia dikarenakan pencatatan data penjualan yang masih dilakukan secara manual. Akibatnya, terdapat produk yang kurang laris dalam persediaan dan produk yang habis. Nilai *centroid* c1 adalah 64,7% dari produk yang paling laris, *centroid* c2 adalah 24,3% dari produk yang cukup laris, dan *centroid* c3 adalah 24,3% dari produk yang cukup laris. Semua ini

dihasilkan dengan menggunakan algoritma *K-Means Clustering*. *Centroid* dengan nilai tertinggi, yaitu *c1* sebesar 11,1%, mengindikasikan produk yang kurang laris, sedangkan *centroid c2* sebesar 24,3% mengindikasikan produk yang cukup laris. Data ini dapat digunakan oleh manajemen untuk mengelola persediaan produk secara lebih efektif.

Untuk mencapai tujuan dari penelitian ini, penulis akan memanfaatkan sumber data resmi dari Susenas (Survei Sosial Ekonomi Nasional) dan BPS (Badan Pusat Statistik Indonesia, 2024). Sebagai tambahan, penulis juga memberikan dasar teoritis yang kuat dan mendukung argumen dengan mengacu pada literatur dan sumber informasi lainnya yang relevan dalam penelitian ini. Selain itu, Penelitian ini akan melihat serta mempertimbangkan masalah dan kekurangan yang mungkin muncul selama pengumpulan dan analisis data. Penelitian ini juga akan membahas metode yang mungkin digunakan untuk melakukan validasi hasil penelitian melalui metode yang relevan (Thoriq Basalamah & Setyadi, 2023).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, masalah yang dapat dirumuskan dalam penelitian ini adalah:

1. Bagaimana cara menggunakan algoritma *K-Means Clustering* untuk menganalisis tingkat penyelesaian pendidikan di Indonesia?
2. Apa saja langkah-langkah yang harus dilakukan untuk menyiapkan data pendidikan untuk analisis *clustering RapidMiner*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan algoritma *K-Means Clustering* menggunakan *RapidMiner* untuk menganalisis data tingkat penyelesaian pendidikan di Indonesia.
2. Memberikan rekomendasi yang dapat membantu pembuat kebijakan dalam meningkatkan kualitas pendidikan dan aksesibilitas di Indonesia.

1.4 Manfaat Penelitian

Manfaat dalam penelitian ini adalah:

1. Hasil dari proses *clustering* dapat dimanfaatkan untuk memberikan rekomendasi kebijakan yang signifikan dan lebih tepat dengan sasaran serta berbasis data. Misalnya, daerah-daerah dengan tingkat pendidikan rendah dapat menjadi fokus intervensi kebijakan.
2. Penelitian ini dapat menjadi acuan bagi penelitian-penelitian berikutnya yang memanfaatkan teknik *data mining* dalam analisis data pendidikan, sehingga berkontribusi pada pengembangan ilmu pengetahuan di bidang ini.

1.5 Batasan Penelitian

Batasan yang ada dalam penelitian ini adalah:

1. Penelitian ini terbatas, memanfaatkan data tingkat penyelesaian pendidikan yang tersedia dan dapat diakses secara publik. Data tersebut mungkin terbatas pada periode waktu tertentu dan tidak mencakup semua jenjang pendidikan atau provinsi secara lengkap.
2. Penelitian ini fokus pada variabel-variabel tertentu seperti tingkat penyelesaian pendidikan menurut jenjang pendidikan dan provinsi. Variabel lain yang mungkin relevan, seperti latar belakang sosial ekonomi, tidak termasuk dalam analisis.